



Universidade de Vigo

Pablo Romero Fresco
Campus Universitario Lagoas-Marcosende
36200 Vigo, Spain
promero@uvigo.es

13 October 2019

Before the Federal Communications Commission

In the Matter of Petition for Declaratory Ruling and Petition for Rulemaking on Live Closed Captioning Quality Metrics and the Use of Automated Speech Recognition Technologies

My name is Pablo Romero-Fresco. I'm a researcher in media accessibility at the Universidade de Vigo (Spain) and Honorary Professor in Translation and Filmmaking at the University of Roehampton (London). For the past 15 years, and along with colleagues from my research group GALMA, I've been doing research on live captions in different countries. For this purpose, I developed the NER model, a method to assess the quality of live captions that is now being used by governmental regulators, broadcasters and/or companies in countries such as Spain, the UK, France, Italy, Germany Holland, Belgium, Switzerland, Poland, Austria, Finland, Australia, South Africa, Brasil, Canada and now the US.

The governmental regulators with whom we have collaborated have so far mainly chosen between soft and hands-on approaches. Soft approaches often involve the adoption of live captioning quality criteria in official guidelines (captions must be accurate, complete, synchronous with the audio, etc.) and the inclusion of mechanisms to deal with user complaints. They have proved useful to raise awareness and trigger discussions amongst key stakeholders. However, many viewers do not file complaints and the quality criteria (accuracy, completeness, etc.) are often too general to be used in a comparable and consistent manner. This means that soft approaches may end up having less impact on the quality of the captions and the experience of the viewers than initially desired.

In contrast, regulators in countries such as the UK and Canada have opted for hands-on approaches. In these countries, [Ofcom](#) and the [CRTC](#) have decided to officially adopt the NER model to analyse the quality of selected samples of live captions from different TV genres (news, chat shows, sports, etc.). This has enabled these regulators to obtain comparable results of live captioning quality across companies and broadcasters.

Some of the quality evaluation systems currently in use in the US (such as the Accuracy Readability Rating) are very useful to compare live captions to the original soundtrack of a programme and determine what is missing and/or altered in the captions. However, they are often based on words, rather than on meaning, and they do not always take into account the impact that captioning errors have on viewers' comprehension. Thus, errors involving different types of words (nouns, verbs, etc.) will be scored -1 or -0.5 regardless of the effect they have on the viewers' comprehension in the context in which they occur.

In contrast, the NER model adopts a more user-centric approach, with three degrees of severity:

- Minor errors (-0.25) : the error has little or no impact on the viewers' comprehension ("That was a great goal by *a* Ryan Giggs").
- Standard errors (-0.5): the error causes confusion and the loss of information ("He's a buy you a bull asset" instead of "He's a valuable asset").
- Serious errors: the error introduces a new meaning that is credible in the context in which it occurs ("Government funding for universities has been cut by 15%", instead of "Government funding for universities has been cut by 50%"). Some viewers with hearing loss refer to these errors as "lies".

In other words, what is important here is not the type of word that is involved in the error, but the impact that it has on viewers' comprehension.

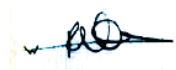
Undoubtedly, there is a degree of subjectivity involved in deciding between the different types of errors. This is why we have set up a short training system for NER evaluators. In the official assessment set up in the UK, for instance, the average discrepancy between the assessments of the different evaluators was 0.19%, that is, the equivalent of 0.5 in a 0-10 scale, which is virtually negligible. It is possible to produce a fully automated system that can assess the quality of live captions, but the ones we have seen so far have been very word-based and have not been able to account for the impact that captioning errors have on the viewers' experience. After all, if in many cases unsupervised, fully-automatic live captions are still not up to scratch, assessing their quality without human assistance may be an even bigger challenge.

So far, the NER model has been used with different captioning methods, such as standard and specialised keyboards (Velotype, in Holland), stenography (in Canada), speech recognition /realtime voice writing (in most European countries) and even unsupervised automatic captions (UK). The results obtained have proved to be aligned with the viewers' opinions of captioning quality in different countries. The model provides results regarding accuracy rate but also an overall assessment of quality, including other important aspects such as delay or speed of captions. It also provides captioners with detailed feedback so they can learn what did not work and how to solve it. Indeed, most of the captioning companies and broadcasters that have used the model for quality assessment have also incorporated it in captioning training. This has helped captioners to be more aware of how their captions are likely to be received by the viewers and crucially, as in the case of the UK, it has led to a [significant increase](#) in the quality of live captioning in the country.

To conclude, since captions are intended to give an ever-increasing number of viewers full access to audiovisual content, the NER model strives to put those viewers at the center, by encouraging captioners to consider the impact that their captions may have on the viewers' comprehension and by assessing quality on the basis of the viewers' experience.

We welcome the FCC's intention to consider the adoption of metrics to assess live captioning, as it is likely to lead to an improvement in captioning quality and, ultimately, in the viewers' experience. Regardless of whether the NER model is considered or not, we are happy to help and to support the adoption of any model that can account for this.

Respectfully Submitted,

A handwritten signature in blue ink, appearing to read 'P. Romero Fresco', with a stylized flourish at the end.

Dr. Pablo Romero Fresco
Ramón y Cajal Researcher, Universidade de Vigo (Spain)
Honorary Professor of Translation and Filmmaking, University of Roehampton (UK)
Director de [GALMA](#) (Galician Observatory for Media Accessibility)